Chapter 222

Group-Sequential Superiority by a Margin Tests for the Difference of Two Proportions (Simulation) (Legacy)

This procedure uses simulation for the calculation of the boundaries as well as for calculation of power (and sample size). Futility boundaries are limited. A variety of test statistics are available.

Introduction

This procedure can be used to determine power, sample size and/or boundaries for group sequential superiority by a margin tests comparing the difference between proportions from two groups. These tests are sometimes referred to as non-zero (or non-unity, for ratios and odds ratios) null tests. The tests that can be simulated in this procedure are the common two-sample Z-test with or without pooled standard error and with or without continuity correction, the T-test, and three score tests. Significance and futility boundaries can be produced. The spacing of the looks can be equal or custom specified. Boundaries can be computed based on popular alpha- and beta-spending functions (O'Brien-Fleming, Pocock, Hwang-Shih-DeCani Gamma family, linear) or custom spending functions. Boundaries can also be input directly to verify alpha- and/or beta-spending properties. Futility boundaries can be binding or non-binding. Maximum and average (expected) sample sizes are reported as well as the alpha and/or beta spent and incremental power at each look. Corresponding P-Value boundaries are given for each boundary statistic. Plots of boundaries are also produced.

Technical Details

This section outlines many of the technical details of the techniques used in this procedure including the simulation summary, the test statistic details, and the use of spending functions.

An excellent text for the background and details of many group-sequential methods is Jennison and Turnbull (2000).

Simulation Procedure

In this procedure, a large number of simulations are used to calculate boundaries and power using the following steps

1. Based on the specified proportions, random samples of size N1 and N2 are generated under the null distribution and under the alternative distribution. These are simulated samples as though the final look is reached.
2. For each sample, test statistics for each look are produced. For example, if N1 and N2 are 100 and there are 5 equally spaced looks, test statistics are generated from the random samples at N1 = N2 = 20, N1 = N2 = 40, N1 = N2 = 60, N1 = N2 = 80, and N1 = N2 = 100 for both null and alternative samples.

3. To generate the first significance boundary, the null distribution statistics of the first look (e.g., at N1 = N2 = 20) are ordered and the percent of alpha to be spent at the first look is determined (using either the alpha-spending function or the input value). The statistic for which the percent of statistics above (or below, as the case may be) that value is equal to the percent of alpha to be spent at the first look is the boundary statistic. It is seen here how important a large number of simulations is to the precision of the boundary estimates.
4. All null distribution samples that are outside the first significance boundary at the first look are removed from consideration for the second look. If binding futility boundaries are also being computed, all null distribution samples with statistics that are outside the first futility boundary are also removed from consideration for the second look. If non-binding futility boundaries are being computed, null distribution samples with statistics outside the first futility boundary are not removed.
5. To generate the second significance boundary, the remaining null distribution statistics of the second look (e.g., at N1 = N2 = 40) are ordered and the percent of alpha to be spent at the second look is determined (again, using either the alpha-spending function or the input value). The percent of alpha to be spent at the second look is multiplied by the total number of simulations to determine the number of the statistic that is to be the second boundary statistic. The statistic for which that number of statistics is above it (or below, as the case may be) is the second boundary statistic. For example, suppose there are initially 1000 simulated samples, with 10 removed at the first look (from, say, alpha spent at Look 1 equal to 0.01), leaving 990 samples considered for the second look. Suppose further that the alpha to be spent at the second look is 0.02. This is multiplied by 1000 to give 20. The 990 still-considered statistics are ordered and the 970th (20 in from 990) statistic is the second boundary.
6. All null distribution samples that are outside the second significance boundary and the second futility boundary, if binding, at the second look are removed from consideration for the third look (e.g., leaving 970 statistics computed at N1 = N2 = 60 to be considered at the third look). Steps 4 and 5 are repeated until the final look is reached.

Futility boundaries are computed in a similar manner using the desired beta-spending function or custom beta-spending values and the alternative hypothesis simulated statistics at each look. For both binding and non-binding futility boundaries, samples for which alternative hypothesis statistics are outside either the significance or futility boundaries of the previous look are excluded from current and future looks.

Because the final futility and significance boundaries are required to be the same, futility boundaries are computed beginning at a small value of beta (e.g., 0.0001) and incrementing beta by that amount until the futility and significance boundaries meet.

When boundaries are entered directly, this procedure uses the null hypothesis and alternative hypothesis simulations to determine the number of test statistics that are outside the boundaries at each look. The cumulative proportion of alternative hypothesis statistics that are outside the significance boundaries is the overall power of the study.

Small Sample Considerations

When the sample size is small, say 200 or fewer per group, the discrete nature of the number of possible differences in proportions in the sampling distribution comes into play. This has led to a large number of proposed tests for comparing two proportions (or testing the 2 by 2 table of counts). For example, Upton (1982) considers twenty-two alternative tests for comparing two proportions. Sweeping statements about the power of one test over another are impossible to make, because the size of the Type I error depends upon the proportions used. At some proportions, some tests are overly conservative while others are not, while at other proportions the reverse may be true.

This simulation procedure, however, is based primarily on the ordering of the sample statistics in the simulation. The boundaries are determined by the spending function alphas. Thus, if a test used happens to be conservative in the single-look traditional sense, the boundaries chosen in the simulation results of this procedure will generally remove the conservative nature of the test. This makes comparisons to the one-look case surprising in many cases.

Definitions

Suppose you have two populations from which dichotomous (binary) responses will be recorded. The probability (or risk) of obtaining the event of interest in population 1 (the treatment group) is p_1 and in population 2 (the control group) is p_2. The corresponding failure proportions are given by $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$.

The assumption is made that the responses from each group follow a binomial distribution. This means that the event probability, p_i, is the same for all subjects within the group and that the response from one subject is independent of that of any other subject.

Random samples of *m* and *n* individuals are obtained from these two populations. The data from these samples can be displayed in a 2-by-2 contingency table as follows

Group	Success	Failure	Total
Treatment	*a*	*c*	*m*
Control	*b*	*d*	*n*
Total	*s*	*f*	*N*

The following alternative notation is also used.

Group	Success	Failure	Total
Treatment	x_{11}	x_{12}	n_1
Control	x_{21}	x_{22}	n_2
Total	m_1	m_2	N

The binomial proportions p_1 and p_2 are estimated from these data using the formulae

$$\hat{p}_1 = \frac{a}{m} = \frac{x_{11}}{n_1} \text{ and } \hat{p}_2 = \frac{b}{n} = \frac{x_{21}}{n_2}$$

Comparing Two Proportions

Let $p_{1.0}$ represent the group 1 proportion tested by the null hypothesis, H_0. The power of a test is computed at a specific value of the proportion which we will call $p_{1.1}$. Let δ represent the smallest difference (margin of equivalence) between the two proportions that still results in the conclusion that the new treatment is not inferior to the current treatment. For a superiority test, $\delta < 0$. The set of statistical hypotheses that are tested is

$$\mathrm{H}_0: p_{1.0} - p_2 \leq \delta \text{ versus } \mathrm{H}_1: p_{1.0} - p_2 > \delta$$

which can be rearranged to give

$$\mathrm{H}_0: p_{1.0} \leq p_2 + \delta \text{ versus } \mathrm{H}_1: p_{1.0} > p_2 + \delta$$

There are multiple methods of specifying the margin of superiority. The most direct is to simply give values for p_2 and $p_{1.0}$. However, it is often more meaningful to give p_2 and then specify $p_{1.0}$ implicitly by specifying the difference, ratio, or odds ratio. Mathematically, assuming higher proportions are better, the definitions of these parameterizations are

Parameter	**Computation**	**Hypotheses**
Difference	$\delta = p_{1.0} - p_2$	$\mathrm{H}_0: p_{1.0} - p_2 \leq \delta_0$ vs. $\mathrm{H}_1: p_{1.0} - p_2 > \delta_0, \quad \delta_0 < 0$
Ratio	$\phi = p_{1.0} / p_2$	$\mathrm{H}_0: p_1 / p_2 \leq \phi_0$ vs. $\mathrm{H}_1: p_1 / p_2 > \phi_0, \quad \phi_0 < 1$
Odds Ratio	$\psi = Odds_{1.0} / Odds_2$	$\mathrm{H}_0: o_{1.0} / o_2 \leq \psi_0$ vs. $\mathrm{H}_1: o_{1.0} / o_2 > \psi_0, \quad \psi_0 < 1$

Difference

The difference is perhaps the most direct method of comparison between two proportions. It is easy to interpret and communicate. It gives the absolute impact of the treatment. However, there are subtle difficulties that can arise with its interpretation.

One difficulty arises when the event of interest is rare. If a difference of 0.001 occurs when the baseline probability is 0.40, it would be dismissed as being trivial. However, if the baseline probably of a disease is 0.002, a 0.001 decrease would represent a reduction of 50%. Thus, interpretation of the difference depends on the baseline probability of the event.

The following example is used to convey the concept of a *superiority* test. Suppose 60% of patients respond to the current treatment method ($p_2 = 0.60$). If the response rate of the new treatment is no less than 5 percentage points better ($\delta = 0.05$) than the existing treatment, it will be considered to be superior. Substituting these figures into the statistical hypotheses gives

$$H_0: \delta \leq 0.05 \text{ versus } H_1: \delta > 0.05$$

In this example, when the null hypothesis is rejected, the concluded alternative is that the response rate is at least 65%, which means that the new treatment is superior to the current treatment.

Test Statistics

This section describes the test statistics that are available in this procedure.

Z Test (Pooled and Unpooled)

This test statistic was first proposed by Karl Pearson in 1900. Although this test can be expressed as a Chi-Square statistic, it is expressed here as a *z* so that it can be used for one-sided hypothesis testing.

Both *pooled* and *unpooled* versions of this test have been discussed in the statistical literature. The pooling refers to the way in which the standard error is estimated. In the pooled version, the two proportions are averaged, and only one proportion is used to estimate the standard error. In the unpooled version, the two proportions are used separately.

The formula for the test statistic is

$$z_t = \frac{\hat{p}_1 - \hat{p}_2}{\hat{\sigma}_D}$$

Pooled Version

$$\hat{\sigma}_D = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

Unpooled Version

$$\hat{\sigma}_D = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Continuity Correction

Frank Yates is credited with proposing a correction to the Pearson Chi-Square test for the lack of continuity in the binomial distribution. However, the correction was in common use when he proposed it in 1922.

The continuity corrected z-test is

$$z = \frac{(\hat{p}_1 - \hat{p}_2) + \frac{F}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\hat{\sigma}_D}$$

where *F* is -1 for upper-tailed, 1 for lower-tailed, and either -1 or 1 for two-sided hypotheses, depending on whether the numerator difference is positive or negative.

T-Test

Based on a study of the behavior of several tests, D'Agostino (1988) and Upton (1982) proposed using the usual two-sample *t*-test for testing whether two proportions are equal. One substitutes a '1' for a success and a '0' for a failure in the usual, two-sample *t*-test formula. The test statistic is computed as

$$t_{N-2} = (ad - bc)\left(\frac{N-2}{N(nac + mbd)}\right)^{\frac{1}{2}}$$

which can be compared to the *t* distribution with *N*-2 degrees of freedom.

Miettinen and Nurminen's Likelihood Score Test

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the difference is equal to a specified, non-zero, value, δ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 - \tilde{p}_2 = \delta_0$, are used in the denominator. A correction factor of *N*/(*N*-1) is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic. The formula for computing this test statistic is

$$z_{MND} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\hat{\sigma}_{MND}}$$

where

$$\hat{\sigma}_{MND} = \sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)\left(\frac{N}{N-1}\right)}$$

$$\tilde{p}_1 = \tilde{p}_2 + \delta_0$$

$$\tilde{p}_1 = 2B\cos(A) - \frac{L_2}{3L_3}$$

$$A = \frac{1}{3}\left[\pi + \cos^{-1}\left(\frac{C}{B^3}\right)\right]$$

$$B = \operatorname{sign}(C)\sqrt{\frac{L_2^2}{9L_3} - \frac{L_1}{3L_3}}$$

$$C = \frac{L_2^3}{27L_3^3} - \frac{L_1L_2}{6L_3^2} + \frac{L_0}{2L_3}$$

$$L_0 = x_{21}\delta_0(1-\delta_0)$$

$$L_1 = [n_2\delta_0 - N - 2x_{21}]\delta_0 + m_1$$

$$L_2 = (N + n_2)\delta_0 - N - m_1$$

$$L_3 = N$$

Farrington and Manning's Likelihood Score Test

Farrington and Manning (1990) proposed a test statistic for testing whether the difference is equal to a specified value δ_0. The regular MLE's, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 - \tilde{p}_2 = \delta_0$, are used in the denominator. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{FMD} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)}}$$

where the estimates $\tilde{p}_1$ and $\tilde{p}_2$ are computed as in the corresponding test of Miettinen and Nurminen (1985) given above.

Gart and Nam's Likelihood Score Test

Gart and Nam (1990), page 638, proposed a modification to the Farrington and Manning (1988) difference test that corrects for skewness. Let $z_{FMD}(\delta)$ stand for the Farrington and Manning difference test statistic described above. The skewness corrected test statistic, z_{GND}, is the appropriate solution to the quadratic equation

$$(-\tilde{\gamma})z_{GND}^2 + (-1)z_{GND} + (z_{FMD}(\delta) + \tilde{\gamma}) = 0$$

where

$$\tilde{\gamma} = \frac{\tilde{V}^{3/2}(\delta)}{6}\left(\frac{\tilde{p}_1\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{n_1^2} - \frac{\tilde{p}_2\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{n_2^2}\right)$$

Spending Functions

Spending functions can be used in this procedure to specify the proportion of alpha or beta that is spent at each look without having to specify the proportion directly.

Spending functions have the characteristics that they are increasing and that

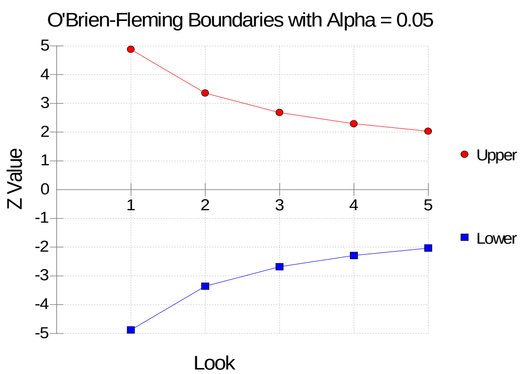
$$\alpha(0) = 0$$

$$\alpha(1) = \alpha$$

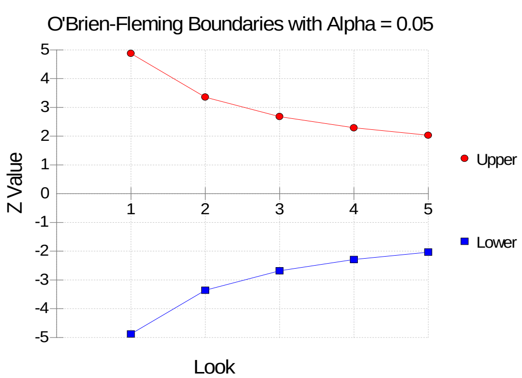
The last characteristic guarantees a fixed α level when the trial is complete. This methodology is very flexible since neither the times nor the number of analyses must be specified in advance. Only the functional form of $\alpha(\tau)$ must be specified.

PASS provides several popular spending functions plus the ability to enter and analyze your own percents of alpha or beta spent. These are calculated as follows (beta may be substituted for alpha for beta-spending functions):

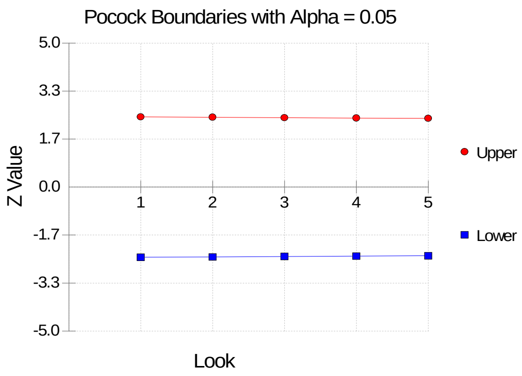
1. Hwang-Shih-DeCani (gamma family) $\alpha\left[\frac{1-e^{-\gamma t}}{1-e^{-\gamma}}\right], \gamma \neq 0; \quad \alpha t, \gamma = 0$



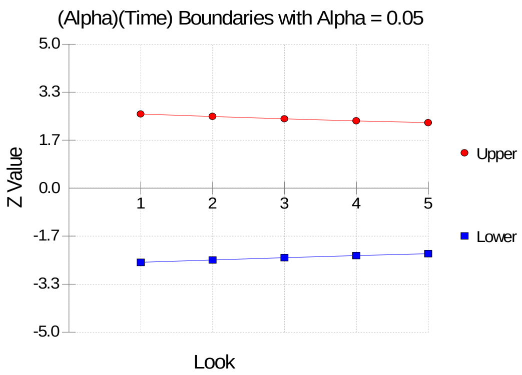
2. O'Brien-Fleming Analog $2 - 2\Phi\left(\frac{Z_{\alpha/2}}{\sqrt{t}}\right)$



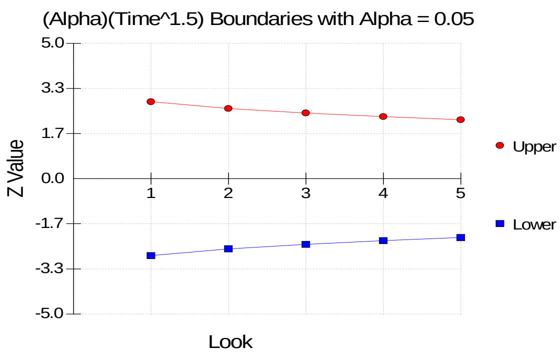
3. Pocock Analog $\alpha \cdot \ln(1 + (e - 1)t)$



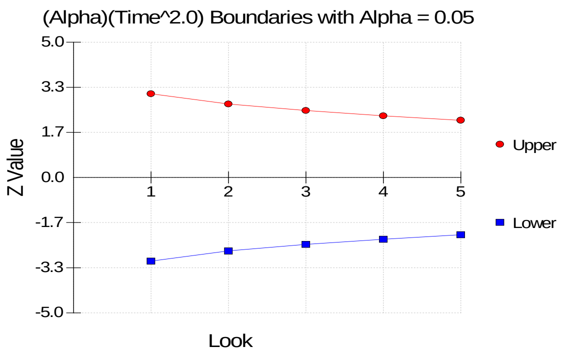
4. Alpha * time $\alpha \cdot t$



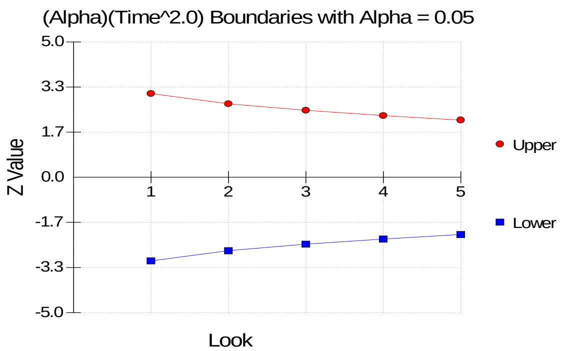
5. Alpha * time^1.5 $\alpha \cdot t^{3/2}$



6. Alpha * time^2 $\alpha \cdot t^2$



7. Alpha * time^C $\alpha \cdot t^C$



8. User Supplied Percents

A custom set of percents of alpha to be spent at each look may be input directly.

The O'Brien-Fleming Analog spends very little alpha or beta at the beginning and much more at the final looks. The Pocock Analog and (Alpha or Beta)(Time) spending functions spend alpha or beta more evenly across the looks. The Hwang-Shih-DeCani (C) (gamma family) spending functions and (Alpha or Beta)(Time^C) spending functions are flexible spending functions that can be used to spend more alpha or beta early or late or evenly, depending on the choice of C.

Example 1 – Power and Output

A clinical trial is to be conducted over a two-year period to compare the proportion response of a new treatment to that of the current treatment. The current response proportion is 0.58. The researchers would like to show that the new treatment is at least 0.05 better than the standard treatment. Although the researchers do not know the true proportion of patients that will survive with the new treatment, they would like to examine the power that is achieved if the proportion under the new treatment is 0.68. The sample size at the final look is to be 1000 per group. Testing will be done at the 0.05 significance level. A total of five tests are going to be performed on the data as they are obtained. The O'Brien-Fleming (Analog) boundaries will be used.

Find the power and test boundaries assuming equal sample sizes per arm.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For .... **Power**
Test Type.. **Z-Test (Pooled)**
Higher Proportions Are .... **Better**
Simulations **100000**
Random Seed... **3208770** (for Reproducibility)
Alpha... **0.05**
Group Allocation .... **Equal (N1 = N2)**
Sample Size Per Group **1000**
Input Type... **Proportions**
P1.0 (Group 1 Proportion|H0) **0.63**
P1.1 (Group 1 Proportion|H1) **0.68**
P2 (Group 2 Proportion)................................ **0.58**

Looks & Boundaries Tab

Specification of Looks and Boundaries.......... **Simple**
Number of Equally Spaced Looks.................. **5**
Alpha Spending Function.............................. **O'Brien-Fleming Analog**
Type of Futility Boundary .... **None**

Output

Click the Calculate button to perform the calculations and generate the following output.

Scenario 1 Numeric Results for Group Sequential Testing Proportion Difference = D0

Solve For:	Power
Hypotheses:	H0: Proportion 1 - Proportion 2 = D0; H1: Proportion 1 - Proportion 2 > D0
Test Statistic:	Z-Test (Pooled)
Zero Adjustment Method:	None
Alpha-Spending Function:	O'Brien-Fleming Analog
Beta-Spending Function:	None
Futility Boundary Type:	None
Number of Looks:	5
Simulations:	100000
Random Seed:	3208770 (User-Entered)

Numeric Summary for Scenario 1

Power			Alpha				
Value	95% LCL	95% UCL	Target	Actual	95% LCL	95% UCL	Beta
0.741	0.738	0.743	0.05	0.05	0.049	0.051	0.259

		Average Sample Size								
		Given H0		Given H1						
N1	N2	Grp1	Grp2	Grp1	Grp2	D0	D1	P1.0	P1.1	P2
1000	1000	992	992	808	808	0.05	0.1	0.63	0.68	0.58

Power	The probability of rejecting a false null hypothesis at one of the looks. It is the total proportion of alternative hypothesis simulations that are outside the significance boundaries.
Power 95% LCL and UCL	The lower and upper confidence limits for the power estimate. The width of the interval is based on the number of simulations.
Target Alpha	The user-specified probability of rejecting a true null hypothesis. It is the total alpha spent.
Alpha or Actual Alpha	The alpha level that was actually achieved by the experiment. It is the total proportion of the null hypothesis simulations that are outside the significance boundaries.
Alpha 95% LCL and UCL	The lower and upper confidence limits for the actual alpha estimate. The width of the interval is based on the number of simulations.
Beta	The probability of accepting a false null hypothesis. It is the total proportion of alternative hypothesis simulations that do not cross the significance boundaries.
N1 and N2	The sample sizes of each group if the study reaches the final look.
Average Sample Size Given H0	The average or expected sample sizes of each group if H0 is true. These are based on the proportion of null hypothesis simulations that cross the significance or futility boundaries at each look.
Average Sample Size Given H1	The average or expected sample sizes of each group if H1 is true. These are based on the proportion of alternative hypothesis simulations that cross the significance or futility boundaries at each look.
D0	The superiority difference is the proportion difference between groups (Grp1 - Grp2) assuming the null hypothesis, H0.
D1	The proportion difference between groups (Grp1 - Grp2) assuming the alternative hypothesis, H1.
P1.0	The proportion used in the simulations for Group 1 under H0.
P1.1	The proportion used in the simulations for Group 1 under H1.
P2	The proportion used in the simulations for Group 2 under H0 and H1.

Summary Statements

A group sequential trial with sample sizes of 1000 and 1000 at the final look achieve 74% power to detect a difference of 0.1 between a treatment group proportion of 0.68 and a control group proportion of 0.58 with a superiority difference of 0.05 with an overall Type I error rate (α) of 0.05 using a one-sided Z-Test (Pooled).

Accumulated Information Details for Scenario 1

	Accumulated Information	Accumulated Sample Size		
Look	Percent	Group 1	Group 2	Total
1	20	200	200	400
2	40	400	400	800
3	60	600	600	1200
4	80	800	800	1600
5	100	1000	1000	2000

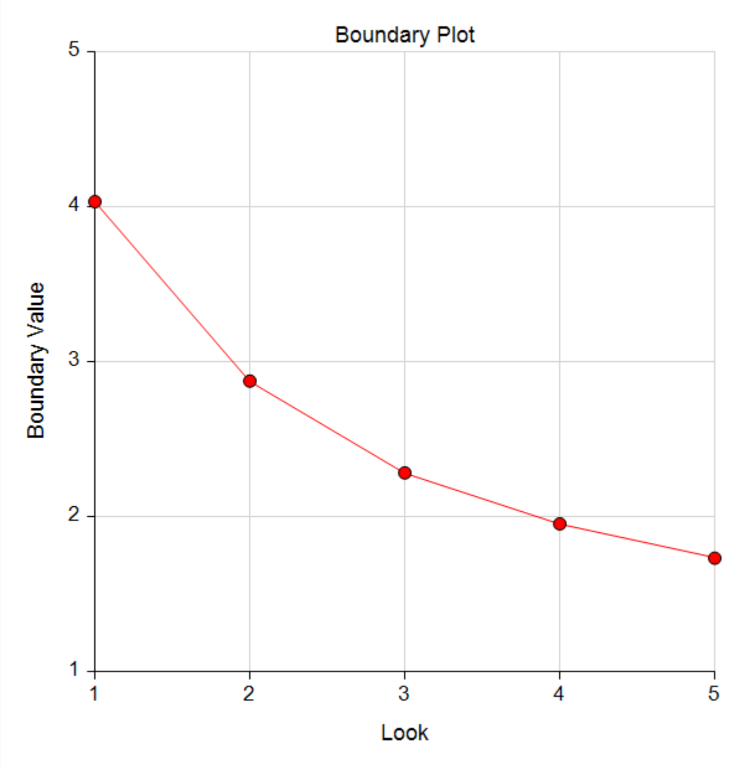
Look — The number of the look.
Accumulated Information Percent — The percent of the sample size accumulated up to the corresponding look.
Accumulated Sample Size Group 1 — The total number of individuals in group 1 at the corresponding look.
Accumulated Sample Size Group 2 — The total number of individuals in group 2 at the corresponding look.
Accumulated Sample Size Total — The total number of individuals in the study (group 1 + group 2) at the corresponding look.

Boundaries for Scenario 1

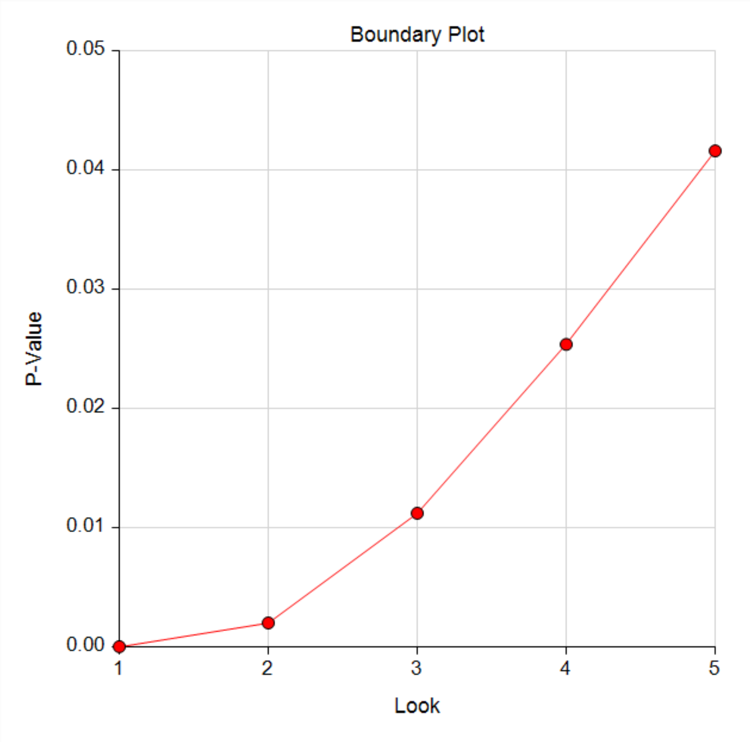
	Significance Boundary	
Look	Z-Value Scale	P-Value Scale
1	4.03334	0.00003
2	2.87539	0.00202
3	2.28296	0.01122
4	1.95353	0.02538
5	1.73232	0.04161

Look — The number of the look.
Significance Boundary Z-Value Scale — The value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.
Significance Boundary P-Value Scale — The value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the Z-Value Boundary and is sometimes called the nominal alpha.

Boundary Plot



Boundary Plot - P-Value



Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1

	Z-Value Boundary			P-Value Boundary		
Look	Value	95% LCL	95% UCL	Value	95% LCL	95% UCL
1	4.03334			0.00003		
2	2.87539	2.82535	2.93753	0.00202	0.00165	0.00236
3	2.28296	2.24903	2.30218	0.01122	0.01066	0.01226
4	1.95353	1.94227	1.98569	0.02538	0.02353	0.02605
5	1.73232	1.70941	1.73971	0.04161	0.04095	0.04369

Look The number of the look.
Z-Value Boundary Value The value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.
P-Value Boundary Value The value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the Z-Value Boundary and is sometimes called the nominal alpha.
95% LCL and UCL The lower and upper confidence limits for the boundary at the given look. The width of the interval is based on the number of simulations.

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

	Signif. Boundary		Target		Actual			
Look	Z-Value Scale	P-Value Scale	Spending Function Alpha	Cum. Spending Function Alpha	Alpha Spent	Cum. Alpha Spent	Proportion H1 Sims Outside Signif. Boundary	Cum. H1 Sims Outside Signif. Boundary
1	4.03334	0.00003	0.000	0.000	0.000	0.000	0.001	0.001
2	2.87539	0.00202	0.002	0.002	0.002	0.002	0.077	0.078
3	2.28296	0.01122	0.009	0.011	0.009	0.011	0.239	0.317
4	1.95353	0.02538	0.017	0.028	0.017	0.028	0.248	0.565
5	1.73232	0.04161	0.022	0.050	0.022	0.050	0.176	0.741

Look The number of the look.
Significance Boundary Z-Value Scale The value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.
Significance Boundary P-Value Scale The value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the Significance Z-Value Boundary and is sometimes called the nominal alpha.
Spending Function Alpha The intended portion of alpha allocated to the particular look based on the alpha-spending function.
Cumulative Spending Function Alpha The intended accumulated alpha allocated to the particular look. It is the sum of the Spending Function Alpha up to the corresponding look.
Alpha Spent The proportion of the null hypothesis simulations resulting in statistics outside the Significance Boundary at this look.
Cumulative Alpha Spent The proportion of the null hypothesis simulations resulting in Significance Boundary termination up to and including this look. It is the sum of the Alpha Spent up to the corresponding look.
Proportion H1 Sims Outside Significance Boundary The proportion of the alternative hypothesis simulations resulting in statistics outside the Significance Boundary at this look. It may be thought of as the incremental power.
Cumulative H1 Sims Outside Significance Boundary The proportion of the alternative hypothesis simulations resulting in Significance Boundary termination up to and including this look. It is the sum of the Proportion H1 Sims Outside Significance Boundary up to the corresponding look.

References

Jennison, C., Turnbull, B.W. 2000. Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall. Boca Raton, FL.
Matsumoto, M. and Nishimura,T. 1998. 'Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator.' ACM Trans. On Modeling and Computer Simulations.
Chow, S.C., Shao, J., and Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
Farrington, C. P. and Manning, G. 1990. 'Test Statistics and Sample Size Formulae for Comparative Binomial Trials with Null Hypothesis of Non-Zero Risk Difference or Non-Unity Relative Risk.' Statistics in Medicine, Vol. 9, pages 1447-1454.
Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.
Gart, John J. and Nam, Jun-mo. 1988. 'Approximate Interval Estimation of the Ratio in Binomial Parameters: A Review and Corrections for Skewness.' Biometrics, Volume 44, Issue 2, 323-338.
Gart, John J. and Nam, Jun-mo. 1990. 'Approximate Interval Estimation of the Difference in Binomial Parameters: Correction for Skewness and Extension to Multiple Tables.' Biometrics, Volume 46, Issue 3, 637-643.
Lachin, John M. 2000. Biostatistical Methods. John Wiley & Sons. New York.
Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, Mass.
Miettinen, O.S. and Nurminen, M. 1985. 'Comparative analysis of two rates.' Statistics in Medicine 4: 213-226.

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 2 – Power with Futility Boundaries

Continuing with Example 1, suppose that the researchers would also like to terminate the study early if there is indication that the treatment is not better than the standard by 0.05.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For	**Power**
Test Type	**Z-Test (Pooled)**
Higher Proportions Are	**Better**
Simulations	**100000**
Random Seed	***Blank* or Random**
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**1000**
Input Type	**Proportions**
P1.0 (Group 1 Proportion|H0)	**0.63**
P1.1 (Group 1 Proportion|H1)	**0.68**
P2 (Group 2 Proportion)	**0.58**

Looks & Boundaries Tab

Specification of Looks and Boundaries	**Simple**
Number of Equally Spaced Looks	**5**
Alpha Spending Function	**O'Brien-Fleming Analog**
Type of Futility Boundary	**Non-Binding**
Number of Skipped Futility Looks	**0**
Beta Spending Function	**O'Brien-Fleming Analog**

Output

Click the Calculate button to perform the calculations and generate the following output.

Scenario 1 Numeric Results for Group Sequential Testing Proportion Difference = D0

Solve For:	Power
Hypotheses:	H0: Proportion 1 - Proportion 2 = D0; H1: Proportion 1 - Proportion 2 > D0
Test Statistic:	Z-Test (Pooled)
Zero Adjustment Method:	None
Alpha-Spending Function:	O'Brien-Fleming Analog
Beta-Spending Function:	O'Brien-Fleming Analog
Futility Boundary Type:	Non-Binding
Number of Looks:	5
Simulations:	100000
Random Seed:	5671104 (Computer-Generated)

Numeric Summary for Scenario 1

Power			Alpha				
Value	95% LCL	95% UCL	Target	Actual	95% LCL	95% UCL	Beta
0.656	0.653	0.659	0.05	0.039	0.038	0.04	0.344

		Average Sample Size								
		Given H0		Given H1						
N1	N2	Grp1	Grp2	Grp1	Grp2	D0	D1	P1.0	P1.1	P2
1000	1000	468	468	677	677	0.05	0.1	0.63	0.68	0.58

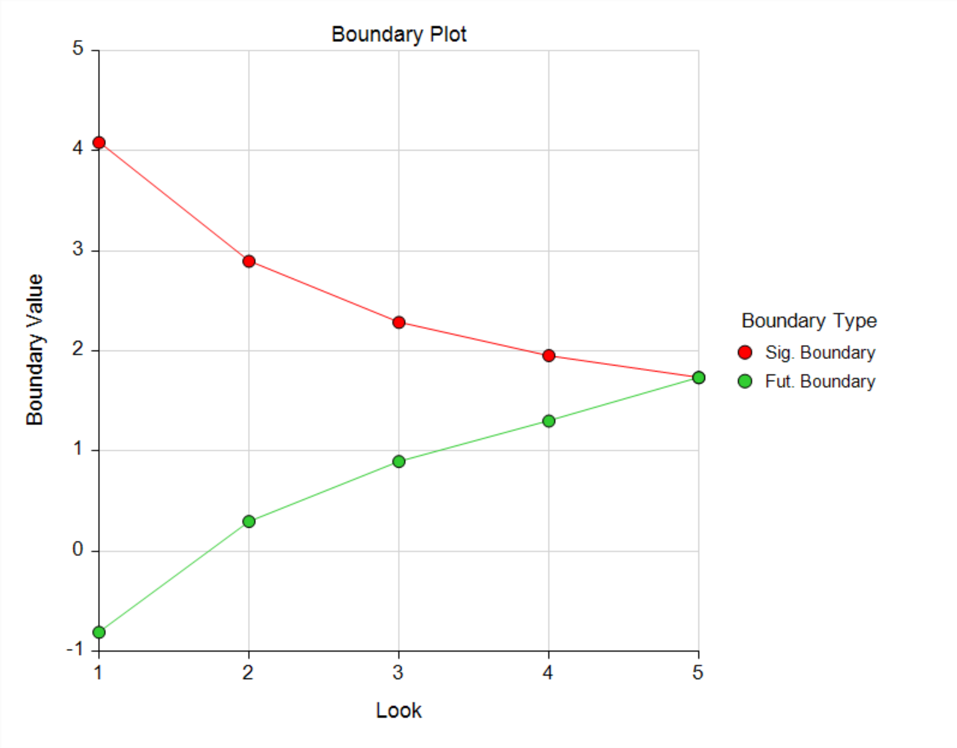
Accumulated Information Details for Scenario 1

	Accumulated	Accumulated Sample Size		
Look	Information Percent	Group 1	Group 2	Total
1	20	200	200	400
2	40	400	400	800
3	60	600	600	1200
4	80	800	800	1600
5	100	1000	1000	2000

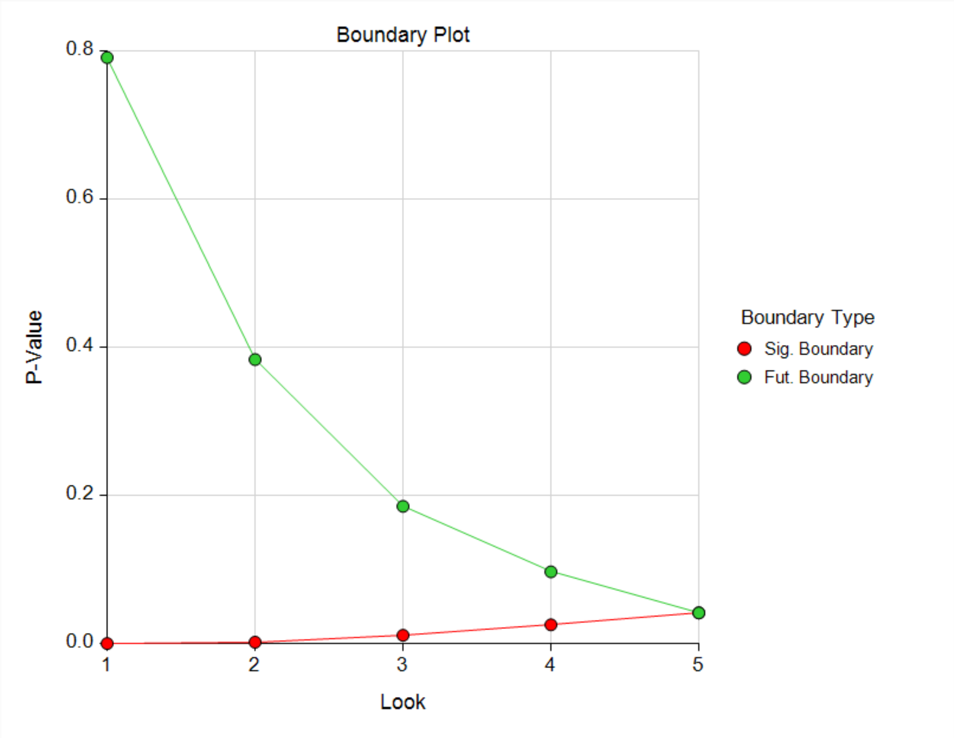
Boundaries for Scenario 1

	Significance Boundary		Futility Boundary	
Look	Z-Value Scale	P-Value Scale	Z-Value Scale	P-Value Scale
1	4.08248	0.00002	-0.81044	0.79116
2	2.89616	0.00189	0.29650	0.38342
3	2.28560	0.01114	0.89563	0.18522
4	1.95353	0.02538	1.29965	0.09686
5	1.73445	0.04142	1.73445	0.04142

Boundary Plot



Boundary Plot - P-Value



Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1

	Z-Value Boundary			P-Value Boundary		
Look	Value	95% LCL	95% UCL	Value	95% LCL	95% UCL
1	4.08248			0.00002		
2	2.89616	2.84627	2.95543	0.00189	0.00156	0.00221
3	2.28560	2.25437	2.30052	0.01114	0.01071	0.01209
4	1.95353	1.94333	1.98473	0.02538	0.02359	0.02599
5	1.73445	1.72451	1.74050	0.04142	0.04089	0.04231

Futility Boundaries with 95% Simulation Confidence Intervals for Scenario 1

	Z-Value Boundary			P-Value Boundary		
Look	Value	95% LCL	95% UCL	Value	95% LCL	95% UCL
1	-0.81044	-0.81825	-0.73755	0.79116		
2	0.29650	0.29508	0.35994	0.38342	0.35945	0.38397
3	0.89563	0.89325	0.89814	0.18522	0.18456	0.18586
4	1.29965	1.29589	1.30459	0.09686	0.09602	0.09751
5	1.73445	1.72311	1.75058	0.04142	0.04001	0.04243

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

	Signif. Boundary		Target		Actual			
Look	Z-Value Scale	P-Value Scale	Spending Function Alpha	Cum. Spending Function Alpha	Alpha Spent	Cum. Alpha Spent	Proportion H0 Sims Outside Futility Boundary	Cum. H0 Sims Outside Futility Boundary
1	4.08248	0.00002	0.000	0.000	0.000	0.000	0.216	0.216
2	2.89616	0.00189	0.002	0.002	0.002	0.002	0.421	0.637
3	2.28560	0.01114	0.009	0.011	0.009	0.011	0.204	0.841
4	1.95353	0.02538	0.017	0.028	0.015	0.027	0.084	0.924
5	1.73445	0.04142	0.022	0.050	0.012	0.039	0.036	0.961

Beta-Spending and Alternative Hypothesis Simulation Details for Scenario 1

	Futility Boundary		Target		Actual			
Look	Z-Value Scale	P-Value Scale	Spending Function Beta	Cum. Spending Function Beta	Beta Spent	Cum. Beta Spent	Proportion H1 Sims Outside Signif. Boundary	Cum. H1 Sims Outside Signif. Boundary
1	-0.81044	0.79116	0.035	0.035	0.035	0.035	0.001	0.001
2	0.29650	0.38342	0.101	0.135	0.100	0.135	0.073	0.074
3	0.89563	0.18522	0.087	0.223	0.087	0.222	0.237	0.311
4	1.29965	0.09686	0.068	0.291	0.068	0.290	0.234	0.545
5	1.73445	0.04142	0.054	0.345	0.054	0.344	0.111	0.656

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 3 – Enter Boundaries

With a set-up similar to Example 2, suppose we wish to investigate the properties of a set of significance (3, 3, 3, 2, 1) and futility (-2, -1, 0, 0, 1) boundaries.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 3** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For ....**Alpha and Power (Enter Boundaries)**
Test Type..**Z-Test (Pooled)**
Higher Proportions Are ....**Better**
Simulations **100000**
Random Seed...***Blank* or Random**
Group Allocation ....**Equal (N1 = N2)**
Sample Size Per Group **1000**
Input Type...**Proportions**
P1.0 (Group 1 Proportion|H0) **0.63**
P1.1 (Group 1 Proportion|H1) **0.68**
P2 (Group 2 Proportion)...............................**0.58**

Looks & Boundaries Tab

Number of Looks **5**
Equally Spaced..**Checked**
Types of Boundaries....................................**Significance and Futility Boundaries**
Significance Boundary ....**3 3 3 2 1** (for looks 1 through 5)
Futility Boundary..**-2 -1 0 0 1** (for looks 1 through 5)

Output

Click the Calculate button to perform the calculations and generate the following output.

Scenario 1 Numeric Results for Group Sequential Testing Proportion Difference = D0

Solve For:	Alpha and Power (Enter Boundaries)
Hypotheses:	H0: Proportion 1 - Proportion 2 = D0; H1: Proportion 1 - Proportion 2 > D0
Test Statistic:	Z-Test (Pooled)
Zero Adjustment Method:	None
Type of Boundaries:	Significance and Futility Boundaries
Number of Looks:	5
Simulations:	100000
Random Seed:	3994733 (Computer-Generated)

Numeric Summary for Scenario 1

Power			Alpha			
Value	95% LCL	95% UCL	Value	95% LCL	95% LCL	Beta
0.895	0.893	0.897	0.155	0.153	0.157	0.105

		Average Sample Size								
		Given H0		Given H1						
N1	N2	Grp1	Grp2	Grp1	Grp2	D0	D1	P1.0	P1.1	P2
1000	1000	738	738	830	830	0.05	0.1	0.63	0.68	0.58

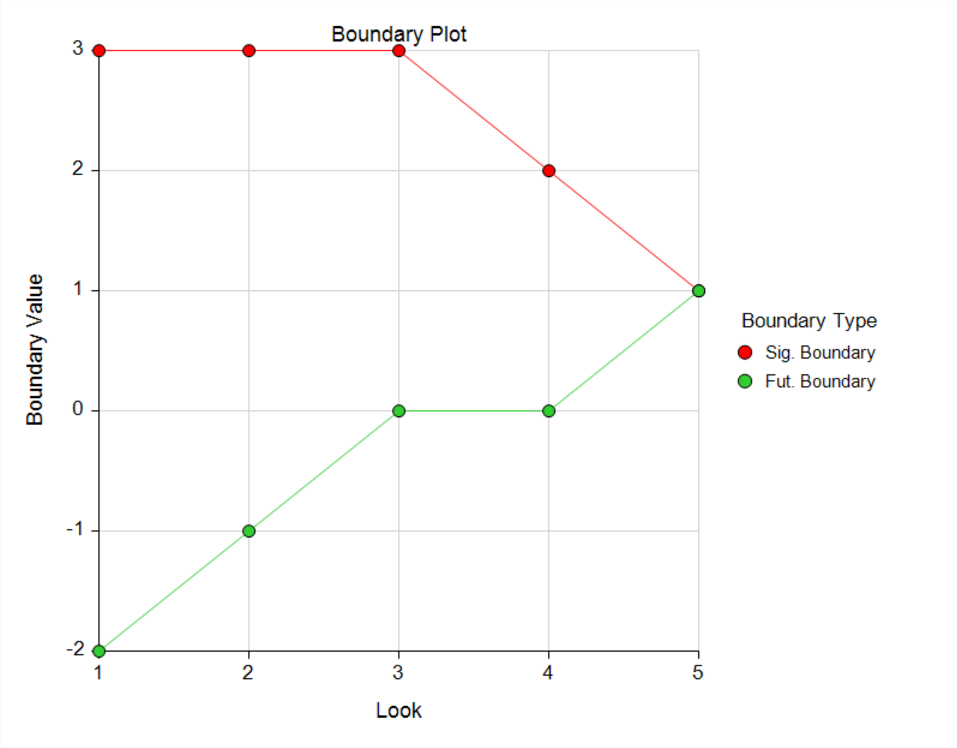
Accumulated Information Details for Scenario 1

	Accumulated Information	Accumulated Sample Size		
Look	Percent	Group 1	Group 2	Total
1	20	200	200	400
2	40	400	400	800
3	60	600	600	1200
4	80	800	800	1600
5	100	1000	1000	2000

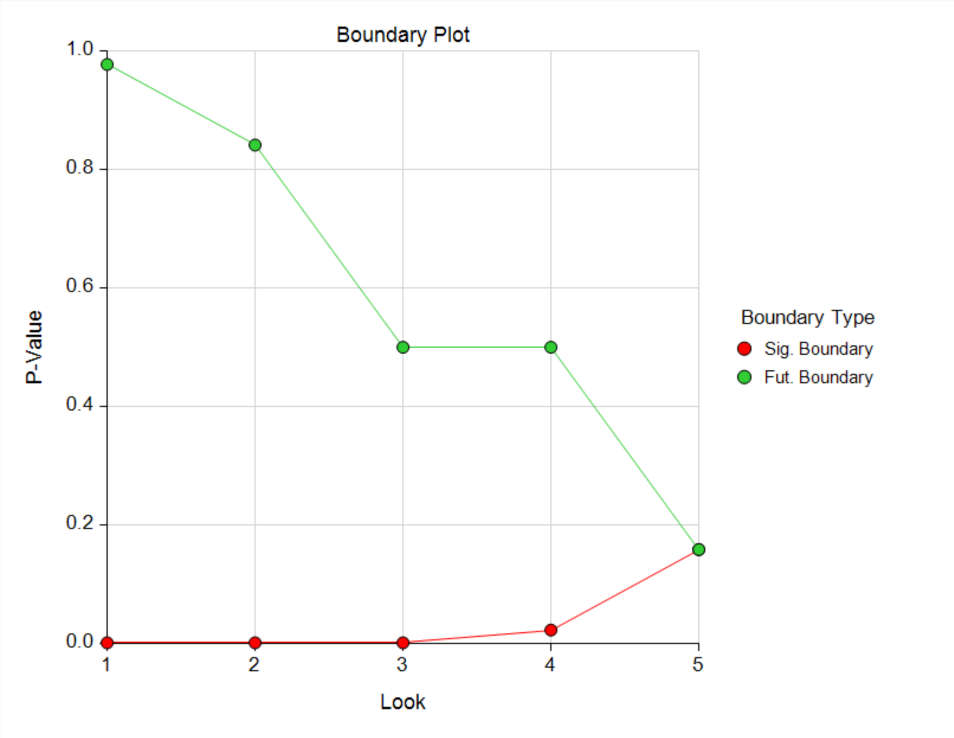
Boundaries for Scenario 1

	Significance Boundary		Futility Boundary	
Look	Z-Value Scale	P-Value Scale	Z-Value Scale	P-Value Scale
1	3	0.00135	-2	0.97725
2	3	0.00135	-1	0.84134
3	3	0.00135	0	0.50000
4	2	0.02275	0	0.50000
5	1	0.15866	1	0.15866

Boundary Plot



Boundary Plot - P-Value



Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

Look	Signif. Boundary Z-Value Scale	Signif. Boundary P-Value Scale	Alpha Spent	Cum. Alpha Spent	Proportion H0 Sims Outside Futility Boundary	Cum. H0 Sims Outside Futility Boundary
1	3	0.00135	0.001	0.001	0.023	0.023
2	3	0.00135	0.001	0.002	0.146	0.169
3	3	0.00135	0.001	0.003	0.335	0.504
4	2	0.02275	0.020	0.023	0.080	0.584
5	1	0.15866	0.132	0.155	0.261	0.845

Beta-Spending and Alternative Hypothesis Simulation Details for Scenario 1

Look	Futility Boundary Z-Value Scale	Futility Boundary P-Value Scale	Beta Spent	Cum. Beta Spent	Proportion H1 Sims Outside Signif. Boundary	Cum. H1 Sims Outside Signif. Boundary
1	-2	0.97725	0.001	0.001	0.023	0.023
2	-1	0.84134	0.006	0.007	0.046	0.069
3	0	0.50000	0.030	0.038	0.065	0.135
4	0	0.50000	0.005	0.043	0.398	0.533
5	1	0.15866	0.062	0.105	0.362	0.895

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 4 – Validation using Simulation

With a set-up similar to Example 2, we examine the power and alpha generated by the set of significance (4.51051, 2.87539, 2.30052, 1.97505, 1.73971) and futility (-0.73502, 0.36264, 0.90165, 1.33501, 1.73971) boundaries.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 4** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For .... **Alpha and Power (Enter Boundaries)**
Test Type .... **Z-Test (Pooled)**
Higher Proportions Are .... **Better**
Simulations **100000**
Random Seed .... ***Blank* or Random**
Group Allocation .... **Equal (N1 = N2)**
Sample Size Per Group **1000**
Input Type .... **Proportions**
P1.0 (Group 1 Proportion|H0) **0.63**
P1.1 (Group 1 Proportion|H1) **0.68**
P2 (Group 2 Proportion) **0.58**

Looks & Boundaries Tab

Number of Looks **5**
Equally Spaced .... **Checked**
Types of Boundaries .... **Significance and Futility Boundaries**
Significance Boundary **4.51051, 2.87539, 2.30052, 1.97505, 1.73971**
Futility Boundary .... **-0.73502, 0.36264, 0.90165, 1.33501, 1.73971**

Output

Click the Calculate button to perform the calculations and generate the following output.

Scenario 1 Numeric Results for Group Sequential Testing Proportion Difference = D0

Solve For:	Alpha and Power (Enter Boundaries)
Hypotheses:	H0: Proportion 1 - Proportion 2 = D0; H1: Proportion 1 - Proportion 2 > D0
Test Statistic:	Z-Test (Pooled)
Zero Adjustment Method:	None
Type of Boundaries:	Significance and Futility Boundaries
Number of Looks:	5
Simulations:	100000
Random Seed:	5030927 (Computer-Generated)

Numeric Summary for Scenario 1

Power			Alpha			
Value	95% LCL	95% UCL	Value	95% LCL	95% LCL	Beta
0.653	0.65	0.655	0.038	0.037	0.04	0.347

		Average Sample Size								
		Given H0		Given H1						
N1	N2	Grp1	Grp2	Grp1	Grp2	D0	D1	P1.0	P1.1	P2
1000	1000	463	463	676	676	0.05	0.1	0.63	0.68	0.58

The values obtained from any given run of this example will vary slightly due to the variation in simulations. The power and alpha generated with these boundaries are very close to the values of Example 2.